AN ARTIFICIAL NEURAL NETWORK FOR PATTERN CLASSIFICATION AND VISUALISATION

SANDRA ONG PI YIN

This project is submitted in partial fulfilment of the requirements for a
Bachelor of Science with Honours
(Cognitive Science)

Faculty of Cognitive Sciences and Human Development
UNIVERSITI MALAYSIA SARAWAK
2010

BORANG PENGESAHAN STATUS TESIS **Gred:**

JUDUL : _____

_____

SESI PENGAJIAN :_____

Saya _____

(HURUF BESAR)

mengaku membenarkan tesis * ini disimpan di Pusat Khidmat Maklumat Akademik, Universiti Malaysia Sarawak dengan syarat-syarat kegunaan seperti berikut:

1. Tesis adalah hakmilik Universiti Malaysia Sarawak.
2. Pusat Khidmat Maklumat Akademik, Universiti Malaysia Sarawak dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Pusat Khidmat Maklumat Akademik, Universiti Malaysia Sarawak dibenarkan membuat pendigitan untuk membangunkan Pangkalan Data Kandungan Tempatan.
4. Pusat Khidmat Maklumat Akademik, Universiti Malaysia Sarawak dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.

** sila tandakan (√)

☐ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan seperti termaktub di dalam AKTA RAHSIA RASMI 1972)

☐ TERHAD (Mengandungi maklumat Terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

☐ TIDAK TERHAD

_____ _____

(TANDATANGAN PENULIS) (TANDATANGAN PENYELIA)

Alamat Tetap:

_____

_____

_____

Tarikh : _____ Tarikh: _____

Catatan:

* Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah, Sarjana dan Sarjana Muda

*Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis ini perlu dikelaskan sebagai TERHAD.

The project entitled ‘An Artificial Neural Network for Pattern Classification and Visualisation’ was prepared by Sandra Ong Pi Yin and submitted to the Faculty of Cognitive Sciences and Human Development in partial fulfillment of the requirements for a Bachelor of Science with Honours (Cognitive Science).

Received for examination by:

(Assoc. Prof. Dr. Teh Chee Siong)

Date:

Grade

Statement of Originality

The work described in this Final Year Project, entitled
"An Artificial Neural Network for Pattern Classification and Visualisation"
is to the best of the author's knowledge that of the author except
where due reference is made.

_____ _____

(Date submitted)

(Student's signature)
Sandra Ong Pi Yin
19809

ACKNOWLEDGMENT

First and foremost, I would like to thank the person behind the motivation for completing this final year project, my supervisor, Dr. Teh Chee Siong. Dr. Teh has been very patient and kind in supervising my every step in order for me to successfully complete this task. I am eternally grateful for his guidance and support. I would also like to extend thanks to Mr. Yii Ming Leong, my laboratory demonstrator for Matlab software. Matlab is the main software used for this project. Mr. Yii has been willing to spare some of his time in teaching me the ins and outs of the software while sharing with me strategies to tackle the technical problems I faced during the completion of the project. Lastly, I'd like to acknowledge my friends and family. I really appreciated the constant support all of you provided throughout the most stressful time of the year. Thank you.

TABLE OF CONTENTS

CHAPTER 3 - SOM TOOLBOX AND INTRODUCTION TO EXPERIMENT

CHAPTER 4 - CLASSIFICATION AND VISUALISATION

CHAPTER 5 - HIGH DIMENSIONAL DATA SET CLASSIFICATION AND VISUALISATION

CHAPTER 6 - CONCLUSION

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

AN ARTIFICIAL NEURAL NETWORK FOR PATTERN CLASSIFICATION AND VISUALISATION

Sandra Ong Pi Yin

The industrial revolution and the birth of computers has led to a deeper exploration of Artificial Neural Network (ANN), where scientist tries to emulate the biological neural network. Today, ANN has proven to be able to imitate the human neural network and perform task such as solving real world problems. This study aims to explore in detail an ANN model that is able to perform the task of pattern classification and visualisation, and as well as to evaluate the performance of this model. This proposed model is know as the Self-Organizing Map (SOM) or Kohonen's Map. In order to determine it's accuracy, the SOM classifier is tested using a few simulated Gaussian data sets and a real world data set, the Pima Indians Diabetes data set. The experiments conducted showed that SOM classifier is able to perform the task of classification and visualisation. However, the classifier obtained a non-impressive accuracy of 73.16% in classifying the real world problem. This results indicate that SOM is not reliable compared to other classification applications in literature and can be enhanced in terms of it's performance.

ABSTRAK

AN ARTIFICIAL NEURAL NETWORK FOR PATTERN CLASSIFICATION AND VISUALISATION

Sandra Ong Pi Yin

Revolusi industri dan penciptaan komputer telah membawa kepada pendalaman dalam pemahaman tentang Rangkaian Neural Buatan. Ahli Sains telah membuktikan bahawa ANN dapat meniru rangkaian neural manusia dan mampu melaksanakan tugas seperti manusia. Projek ini bertujuan untuk mengkaji model ANN yang dapat menjalankan tugas klasifikasi pola serta menilai ketepatan pencapaian model tersebut. Model yang dicadangkan dikenali sebagai Self-Organizing Map (SOM) ataupun Kohonen's Map. SOM diuji dengan menggunakan beberapa set data simulasi Gaussian dan data Pima Indians Diabetes untuk menentukan ketepatan klasifikasinya. Hasil ujikaji menunjukkan SOM mampu menjalankan tugas klasifikasi di samping memberi gambaran pola set data yang digunakan. Akan tetapi, pengklasifikasian mencapai ketepatan hanya sebanyak 73.16% bagi data Pima Indians Diabetes. Dapatan ini menunjukkan bahawa prestasi SOM tidak memuaskan jika berbanding dengan aplikasi pengklasifikasian yang lain.

CHAPTER 1
INTRODUCTION

1.0 Introduction

This chapter briefly introduces the background of the project, the motivation behind the project as well as the objectives and scope of the project.

1.8 Background of Study

Technology advancement began during the industrial revolution beginning in the 1900s. This advancement led to the birth of computers later on in the 1950's. However, the idea of emulating the biological neural networks began before computers came about. In the 1940's, scientist and researchers have already begin exploring the world of Artificial Neural Network (ANN) or also known as Neural Networks (NN). Many theories have formed throughout the years and ANN continued to bloom until the period between 1969 and 1981. The sudden downturn towards ANN was due to media spread of fiction stories regarding the ability of neural network accomplishing many tasks. Therefore, it was not welcomed by the society for they fear robots would trump mankind and dominate the world. However, in the 1982, the idea of neural networks was once again convinced when John Hopfield of Caltech stated that the approach of

Artificial Intelligence should not completely focus on the means of imitating the human neural network but to create machines that are able to solve dynamic problems (Oracle, 2008). Today, ANN has proven to be beneficial to society in performing tasks. Countries around the world have pursued and invested a great deal in the research and development of ANN.

1.9 Artificial Neural Network

Artificial Neural Network (ANN) is the attempt in mimicking the patterns of the human mind (Botts, 1998). According to the Encyclopedia of Educational Technology, the idea of the artificial neural network emerged back in the early 1940's when neuroscientist, Warren McCulloch and mathematician, Walter Pitts hypothesized that "neurons in our nervous system could be regarded as a device for manipulating binary numbers, the computer" (1998). By joining venture, they designed and built a primitive artificial neural network by using electric circuits. Their work has led to the infamous McCulloch-Pitts Theory of Formal Neural Networks (Oracle, 2008).

The next major development in neural network technology was in the year 1959, when Bernard Wildro and Marcian Hoff of the Stanford University USA developed the first neural network that can be applied to real world problems. It is known as the Adaptive Linear Elements (ADALINE) and Multiple Adaptive Linear Elements (MADELINE) (Oracle, 2008).

Advancement in ANN has allowed an increasing number of complex real world problems to be solved. A few of many ANN applications in real life includes classification which consist of pattern and sequence recognition, sequential decision making; data processing which consist of filtering, and clustering; and many more (Wikipedia, 2009).

1.10 Data Classification

What is data? Data are factual information organized for analysis or used for decision making purposes (Answers, 2010). In the field of computers, data refers to a collection of numbers and symbols, represented in a form suiTable to be processed by computers (Himberg, 2004). These data might be obtained by scientific experiments, large databases, or collections of digital documents (Himberg, 2004).

Classification infers the act of classifying. It is a fundamental process that is carried out daily by human beings. Human beings are programmed to have the ability to categorize things so it would be easier to identify or distinguish the differences or the similarity between the groups of things (Modell, 1992). According to Modell (1992), "to classify is to organize or arrange according to class or category". These categories contain data or information that has at least one similarity between them. For example, the common attribute for the class 'fruits' is that it contains seeds. Therefore, 'apple' would be classified as a fruit instead of vegeTable because it fulfills the required attribute. Data classification may also be done according to the importance of that data or how frequent it would be accessed, data content, data size, or even categorized based on who uses the data (TechTarget, 2007). These common attributes or similarity in characteristics or features is known as parameters.

Today, there are many neural network models that have been recognized to perform the task of data classification. These include the simple perceptron model, the Adaptive Linear Element (ADALINE), Multi-Layer Perceptron (MLP), Self-Organizing Map (SOM), the Learning Vector Quantization (LVQ) and more.